

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: April 22, 2004, 13:50:23 ; Search time 18 Seconds  
(without alignments)  
1423.251 Million cell updates/sec

Title: US-09-978-188a-7  
Perfect score: 2527  
Sequence: 1 MYKPPALHWPLIRFLVPL.....TDMPPTEVVDIVEMRENE 492

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	2527	100.0	492 1	ANKH HUMAN
2	2502	99.0	492 1	ANKH MOUSE
3	2489	98.5	492 1	ANKH RAT
4	2317	91.7	492 1	ANKH XENLA
5	2158.5	85.4	501 1	ANKH BRAVE
6	1528.5	60.5	355 1	ANKH TETNG
7	114	4.5	416 1	ANKH KLEOX
8	113.5	4.5	556 1	NU2M PODAN
9	111	4.4	622 1	COX1 BACSU
10	109.5	4.3	610 1	PT2A ARATH
11	107	4.2	422 1	EXTT BACSU
12	107	4.2	533 1	LAT2 RAT
13	106	4.2	334 1	YB77 METUA
14	105	4.2	401 1	LSG1 HAEIN
15	105	4.2	676 1	HPPI METAC
16	103.5	4.1	398 1	BCR HAEIN
17	103	4.1	417 1	LACT ECOLI
18	102.5	4.1	446 1	CITR SALTU
19	102.5	4.1	531 1	LAT2 MOUSE
20	102.5	4.1	533 1	MYIN RHITR
21	102.5	4.1	641 1	NU2M ALLMA
22	101.5	4.0	446 1	CITR SALTU
23	100	4.0	982 1	YS96 CAEEL
24	99	3.9	385 1	Y421 METTH
25	99	3.9	535 1	LAT2 HUMAN
26	98.5	3.9	583 1	NU2M NEUCR
27	98.5	3.9	946 1	YB76 YEAST
28	98	3.9	480 1	YEH4 YEAST
29	98	3.9	532 1	YABW BACST
30	97.5	3.9	346 1	FMLR PONPY
31	97.5	3.9	3803 1	TRAI DROME
32	97	3.8	402 1	YTCB BACSU
33	96.5	3.8	346 1	FMLR GORGO

Sequence Alignment

34	95.5	3.8	416 1	LACT CITR	P47234 citrobacter
35	95.5	3.8	483 1	NORM VIBVU	Q8d9r8 vibrio vuln
36	95	3.8	457 1	NORM ECOS7	P58164 escherichia
37	95	3.8	459 1	NU4M BALPH	P24975 balaeonopter
38	95	3.8	460 1	NU4M ASTPE	P11992 asterina pe
39	95	3.8	464 1	NORM HAEIN	P45272 haemophilus
40	95	3.8	501 1	LXSI CORGL	P35865 corynebacte
41	95	3.8	511 1	ALGB DROME	Q3w3v8 drosophila
42	95	3.8	654 1	NU5M RHIST	P50367 rhizopus st
43	94.5	3.7	507 1	TT12 ARATH	Q91yc3 arabidopsis
44	94.5	3.7	518 1	SP5B BACSU	O00758 bacillus su
45	94	3.7	421 1	VG2 EPIKE	P03660 bacterioph

ALIGNMENTS

RESULT 1  
ANKH HUMAN STANDARD; PRT; 492 AA.  
ID ANKH HUMAN  
AC Q9HCJ1; Q9NCW2; 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Progressive ankylosis protein homolog (ANK).  
GN ANKH OR KIAA1581.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RX MEDLINE=20355194; PubMed=10894769;  
RA Ho A.M., Johnson M.D., Kingsley D.M.;  
RT "Role of the mouse ank gene in control of tissue calcification and  
arthritis.";  
RL Science 289:265-270(2000).  
[2]  
SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20450683; PubMed=10997877;  
RA Nagase T., Kikuno R., Nakayama M., Hirozawa M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes.  
RT XVIII. The complete sequences of 100 new cDNA clones from brain which  
code for large proteins in vitro.";  
RL Mamm. Res. 7:273-281(2000).  
[3]  
SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buettow K.H., Scheffer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldi M.F., Cavaletto T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Pangue C.,  
Raha S.S., Igoe J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Boesch S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield J.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,  
Schnerker A., Schein J.E., Jones S.J.W., Maira M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[4]  
VARIANTS CMDJ SER-375 DEL; PHE-376 DEL AND ALA-380 INS.  
MEDLINE=21313103; PubMed=11326338;

RA Reichenberger E., Tiziani V., Matanabe S., Park L., Ueki Y.,  
 RA Santanna C., Baur S.T., Shiang R., Grange D.K., Beighton P.,  
 RA Gardner J., Hamersma H., Sellars S., Ramesar R., Lital A.C.,  
 RA Sommer A., Raposo do Amaral C.M., Gorlin R.J., Mulliken J.B.,  
 RA Olsen B.R.;  
 RT "Autosomal dominant craniofacial dysplasia is caused by mutations  
 in the transmembrane protein ANK-";  
 RL Am. J. Hum. Genet. 68:1321-1326(2001).  
 RN [5]  
 RP VARIANTS CMDJ ARG-292; ARG-331; SER-375 DEL; PHE-377 DEL; ALA-380 INS  
 RP AND ARG-389.  
 RX MEDLINE=11225282; PubMed=11226272;  
 RA Nienberg P., Thiele H., Chaudler D., Hoehne W., Cunningham M.L.,  
 RA Bitter H., Leschik G., Uhlmann K., Mischung C., Harrop K.,  
 RA Goldblatt J., Borchowitz Z.U., Kotzot D., Westermann F., Mundlos S.,  
 RA Braun H.-S., Laing N., Tinschert S.;  
 RT "Heterozygous mutations in ANKH, the human ortholog of the mouse  
 progressive ankylosis gene, result in craniofacial dysplasia";  
 RL Nat. Genet. 28:37-41(2001).  
 CC -1- FUNCTION: Regulates intra- and extracellular levels of inorganic  
 phosphate (P<sub>pi</sub>), probably functioning as P<sub>pi</sub> transporter.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: Found in osteoblasts from mandibular bone and  
 from iliac bone; not detected in osteoclastic cells.  
 CC -1- DISEASE: Defects in ANKH are the cause of craniofacial dysplasia  
 dysplasia Jackson type (CMDJ) [MIM:123000]. CMDJ is a rare  
 autosomal dominant skeletal disorder characterized by abnormal  
 bone formation and mineralization in membranous as well as  
 endochondral bones. Progressive thickening of the bones can cause  
 narrowing of cranial foramina and can lead to severe visual and  
 neurological impairment, such as facial palsy and deafness.  
 CC -1- SIMILARITY: BELONGS TO THE ANKH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF274753; AAF88039.1; -;  
 DR EMBL; AB046801; BAB13407.1; ALT\_INIT.  
 DR EMBL; BC009835; AAH09835.1; -;  
 DR EMBL; BC014526; AAH14526.1; -;  
 DR Genew; HGNC:15492; ANKH.  
 DR MIM; 605145; -;  
 DR MIM; 123000; -;  
 DR GO; GO:0016021; C:integral to membrane; IDA.  
 DR GO; GO:0019667; C:outer membrane; TAS.  
 DR GO; GO:0030504; F:inorganic diphosphate transporter activity; IDA.  
 DR GO; GO:0005315; F:inorganic phosphate transporter activity; IDA.  
 DR GO; GO:0007665; P:locomotory behavior; NAS.  
 DR GO; GO:0030500; P:regulation of bone mineralization; TAS.  
 DR GO; GO:0001501; P:skeletal development; NAS.  
 KM Transport; Phosphate transport; Transmembrane; Disease mutation;  
 KM Deafness.  
 FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 86 106 POTENTIAL.  
 FT DOMAIN 107 131 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 132 152 POTENTIAL.  
 FT DOMAIN 153 158 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 159 179 POTENTIAL.  
 FT DOMAIN 180 189 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 190 210 POTENTIAL.  
 FT DOMAIN 211 326 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 327 347 POTENTIAL.  
 FT DOMAIN 348 350 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 351 371 POTENTIAL.  
 FT DOMAIN 372 403 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 404 426 POTENTIAL.  
 FT DOMAIN 427 429 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 430 452 POTENTIAL.

FT DOMAIN 453 492 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 493 522 W -> R (in CMDJ).  
 FT VARIANT 292 292 /FTID=VAR 012192.  
 FT VARIANT 331 331 C -> R (in CMDJ).  
 FT VARIANT 331 331 /FTID=VAR 012193.  
 FT VARIANT 375 375 Missing (in CMDJ).  
 FT VARIANT 376 376 /FTID=VAR 012194.  
 FT VARIANT 376 376 Missing (in CMDJ).  
 FT VARIANT 377 377 /FTID=VAR 012195.  
 FT VARIANT 377 377 Missing (in CMDJ).  
 FT VARIANT 380 380 /FTID=VAR 012196.  
 FT VARIANT 380 380 P -> PA (in CMDJ).  
 FT VARIANT 389 389 /FTID=VAR 012197.  
 FT VARIANT 389 389 G -> R (in CMDJ).  
 FT CONFLICT 78 78 /FTID=VAR 012198.  
 FT SEQUENCE 492 AA; 54240 MW; 44BEE3089BDEC6B CRC64;  
 SQ  
 Query Match 100.0%; Score 2527; DB 1; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 7,8e-195;  
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MVEPALTHYMPLEIRFLVPLGITINAIIDFGEQALNGIAVYEDAVEMTASVGLAYSLMK 60  
 DB 1 MVEPALTHYMPLEIRFLVPLGITINAIIDFGEQALNGIAVYEDAVEMTASVGLAYSLMK 60  
 QY 61 FTGPMSPDKNVGLVFNVSKRDRTKAVLCMVVAGALAAVPHLLIAYSDGIYINKLHHY 120  
 DB 61 FTGPMSPDKNVGLVFNVSKRDRTKAVLCMVVAGALAAVPHLLIAYSDGIYINKLHHY 120  
 QY 121 DESVGSKTRAPFLYLAFFPMAMWTHAGILKHKYSFLVGCASISDVIAQVFAVILL 180  
 DB 121 DESVGSKTRAPFLYLAFFPMAMWTHAGILKHKYSFLVGCASISDVIAQVFAVILL 180  
 QY 121 DESVGSKTRAPFLYLAFFPMAMWTHAGILKHKYSFLVGCASISDVIAQVFAVILL 180  
 DB 121 DESVGSKTRAPFLYLAFFPMAMWTHAGILKHKYSFLVGCASISDVIAQVFAVILL 180  
 QY 181 HSHLECRREPLIPILSYMGALVRCITLCLGYKNIHDIIPDRSGELGSDATIRKMLSF 240  
 DB 181 HSHLECRREPLIPILSYMGALVRCITLCLGYKNIHDIIPDRSGELGSDATIRKMLSF 240  
 QY 241 WMLPAILIATQIRISRIIVNLFVSRDLAGSSAAEAVALIATPVGMPEYGMWTEIRAVY 300  
 DB 241 WMLPAILIATQIRISRIIVNLFVSRDLAGSSAAEAVALIATPVGMPEYGMWTEIRAVY 300  
 QY 301 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 DB 301 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 QY 361 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 DB 361 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 QY 361 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 DB 361 PAFDKNPNKLVSTNTVTAHIIKKEFVCNALSTLCPWMPVTPNVSEKILIDIIIGVD 360  
 QY 421 LGVHGAATLGVSLIAGVGSSTVVAIAACVYKOKKKOMNESATGEGDSAMTDMPTTEE 480  
 DB 421 LGVHGAATLGVSLIAGVGSSTVVAIAACVYKOKKKOMNESATGEGDSAMTDMPTTEE 480  
 QY 481 VTDIVEMREENE 492  
 DB 481 VTDIVEMREENE 492  
 RESULT 2  
 ANKH MOUSE STANDARD; PRT; 492 AA.  
 ID ANKH MOUSE  
 AC 09JHZ2; 035138; 035139;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Progressive ankylosis protein (Fn54 protein).  
 OS ANKH OR ANK.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]

D	b		2401	GCGCTGCTTTTCCCGCGCTTTCGGAGGTGCATTGAGAGGATCAGATGGACATCC	2460
Oy			2461	TAACTTGCACTTTTAGTTTTTAACA GTGAAC TGAACCTTTAAGTCTCATCCAGACTTCTPA	2520
D	b		2461	TAACTTGCACTTTTAGTTTTTAACA GTGAAC TGAACCTTTAAGTCTCATCCAGACTTCTPA	2520
Oy			2521	TGCCAGGTGCTGTAGAGGTAACTTTTGAAGTACATATTAACCTGTGTCGTACCTTA	2580
D	b		2521	TGCCAGGTGCTGTAGAGGTAACTTTTGAAGTACATATTAACCTGTGTCGTACCTTA	2580
Oy			2581	GTCAAATCTGCGGCTACAGGTAAATTGAAGATGTACCTAGGACTTCCCTCCACACAT	2640
D	b		2581	GTCAAATCTGCGGCTACAGGTAAATTGAAGATGTACCTAGGACTTCCCTCCACACAT	2640
Oy			2641	ACGATAAAGCAAGACATTTTAACAGATACAGAGTCACTATGTGCTCCTCCGAATA	2700
D	b		2641	ACGATAAAGCAAGACATTTTAACAGATACAGAGTCACTATGTGCTCCTCCGAATA	2700
Oy			2701	AACGATTCGAATCCATGCACTGCACTATATTTTCTAAGTTTGGAAAAGAGGTTTTT	2760
D	b		2701	AACGATTCGAATCCATGCACTGCACTATATTTTCTAAGTTTGGAAAAGAGGTTTTT	2760
Oy			2761	CCTTAAAAAAATTTATAGACAGGCTCACTAAATTGATTTAGTACGAAATCCCTGACCTGA	2820
D	b		2761	CCTTAAAAAAATTTATAGACAGGCTCACTAAATTGATTTAGTACGAAATCCCTGACCTGA	2820
Oy			2821	AAGAACCTPAA CAAAAAAATATTTTAAAGATATAAATATATGCTGATATGTATAT	2880
D	b		2821	AAGAACCTPAA CAAAAAAATATTTTAAAGATATAAATATATGCTGATATGTATAT	2880
Oy			2881	TTATTTTAGGCTATATATCATTTCCATTTTGGCAATTTTCAATATAATGTCTPATACA	2940
D	b		2881	TTATTTTAGGCTATATATCATTTCCATTTTGGCAATTTTCAATATAATGTCTPATACA	2940
Oy			2941	AAAAA 2945	
D	b		2941	AAAAA 2945	
RESULT 3					
LOCUS			ABO46801	3928 bp	mRNA linear PRI 22-FEB-2001
DEFINITION			Homo sapiens mRNA for KIM1581 protein, partial cds.		
ACCESSION			ABO46801		
VERSION			ABO46801.1	GI:10047236	
KEYWORDS			.		
SOURCE			Homo sapiens (human)		
ORGANISM			Homo sapiens		
REFERENCE			Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE			1 (sites)		
			Nagase,T., Kikuno,R., Nakayama,M., Hirokawa,M. and Ohara,O.		
			Prediction of the coding sequences of 100 new cDNA clones from drain		
			XVIII. The complete sequences of 100 new cDNA clones from drain		
			which code for large proteins in vitro		
JOURNAL			DNA Res. 7 (4), 273-281 (2000)		
MEDLINE			20450683		
PUBMED			10997877		
REFERENCE			2 (bases 1 to 3928)		
AUTHORS			Ohara,O., Nagase,T. and Kikuno,R.		
TITLE			Direct Submission		
JOURNAL			Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,		
			Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba		
			293-0812, Japan (E-mail:cdnaif@kazusa.or.jp.		
			URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,		
			Fax:81-438-52-3914)		
FEATURES			Location/Qualifiers		
source			1..3928		
			/organism="Homo sapiens"		
			/mol_type="mRNA"		
			/db_xref="taxon:9606"		
			/clone="fj056930"		
			/note="vector:ipbluescriptII SK plus"		

```

gene
  1..3928
  /gene="KIAA1581"
  <1..1640
  /gene="KIAA1581"
  /note="Start codon is not identified."
  /codon_start=3
  /product="KIAA1581 protein"
  /protein_id="BAB13407.1"
  /db_xref="GI:10047237"
  /translation="APQPRGGAANAAPAEAPPAASVGPAPPAAPAPAPAPAESPRGSRQ
VWQPGTAGTWKPPALTHWPELFLPVLPLGLTVAIIFEGEALNRGIAAIVEDAEMV
ASYGLAASIMKEFTGPMSEDFKNVLEFVLEPDRDETKVLGVNVAIAAAYHTLIVMS
DGLYIINKLHHVEDVSGSKTRAPAFLLAAPPEMDAMTHAGILLHXKISFLVGS
ISDYIAOVVPAIILHSHBERBELPLILSIWGAIVRCTTDLGTYYKINHIDIPR
SGPLGEDATIRKMSFPMPLIALIATORISRPVNFVSDLDGSSAAIEAVALITRA
PYVGHMPEWMLTEIRAVNYPAPDFKNNSNKLVSNSVTAAHKKFIFVCMALSLTLC
FVMPTRPVSEKELIDIGVDFAPAEKCVVPLRFSEPPAVTVRAHITGMILMKTF
FLVAPSSVARIIVLIASLVLPYGVGATLGVSLIAGFVSGSTWAIACAVYRRQ
KKKNENESATFEGEDSAMTDMPPTEVTVDIEMRENE"

ORIGIN

Query Match      99.7%; Score 2936.6; DB 9; Length 3928;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2339; Conservativity 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCTGCCCCCGCTGCGCCCTCGCTCCCGCGAGATGCCCTGCGGGGACAGATGTGTG 60
DB 80 CGCTGCCCCCGCTGCGCCCTCGCTCCCGCGAGATGCCCTGCGGGGACAGATGTGTG 139
QY 61 GGGTACGCCACGCGGGGAGCTATGTGAATTCCGGCGCTCACGACTACTGCGCCCT 120
DB 140 GGGTACGCCACGCGGGGAGCTATGTGAATTCCGGCGCTCACGACTACTGCGCCCT 199
QY 121 GATCCGGTCTTGTTGTCGCCCTCGTGGCATCCAAATAGCCATGCACTTCGGGGAGCAGGC 180
DB 200 GATCCGGTCTTGTTGTCGCCCTCGTGGCATCCAAATAGCCATGCACTTCGGGGAGCAGGC 259
QY 181 CTTGAACCGGGGCAATGCTGCTGCACAGAGAGATGCAGATGCTGGCCACTACGG 240
DB 260 CTTGAACCGGGGCAATGCTGCTGCACAGAGAGATGCAGATGCTGGCCACTACGG 319
QY 241 GCTGGCGTACTCCTCATGAAATTCTTCACGGGTCCCATGATGACTTCAAAAATGTGG 300
DB 320 GCTGGCGTACTCCTCATGAAATTCTTCACGGGTCCCATGATGACTTCAAAAATGTGG 379
QY 301 CCTGGTGTGTGTGAACAGCAAGAGAGAGAGCAAGCAAGCCGTCCTGTGTATGTGGTGGC 360
DB 380 CCTGGTGTGTGTGAACAGCAAGAGAGAGAGCAAGCAAGCCGTCCTGTGTATGTGGTGGC 439
QY 361 AGGGGCGCATGCGTCCGTCCTTTCACACAGCTAGTACTTAATAGATTAGGATACTACAT 420
DB 440 AGGGGCGCATGCGTCCGTCCTTTCACACAGCTAGTACTTAATAGATTAGGATACTACAT 499
QY 421 TATCAATTAATCTGACCAATGTGGAACGAGTGGGTGGGAGCAAGACAGAAAGGCGTTCCT 480
DB 500 TATCAATTAATCTGACCAATGTGGAACGAGTGGGTGGGAGCAAGACAGAAAGGCGTTCCT 559
QY 481 GTACTCGCGCGCTTTCCTTTCATGAGACGCAATGGAGACCATCTGGCAATTCCTT 540
DB 560 GTACTCGCGCGCTTTCCTTTCATGAGACGCAATGGAGACCATCTGGCAATTCCTT 619
QY 541 AAAACACAATATACGTTTCTGTGGGATGTGCTCAATCTAGATGTCAATAGCTACAGT 600
DB 620 AAAACACAATATACGTTTCTGTGGGATGTGCTCAATCTAGATGTCAATAGCTACAGT 679
QY 601 TGTTTTGTAGCCATTTTGTCTTACACATCAACCTGGAATGCGCGGAGGCGCTGCTCATGCC 660
DB 680 TGTTTTGTAGCCATTTTGTCTTACACATCAACCTGGAATGCGCGGAGGCGCTGCTCATGCC 739
QY 661 GATCCTCTCTCTGTGAACAGGGGCGCATTTGTGGCGCTGACACACCCCTGTGCGTGGGACTATCA 720
DB 740 GATCCTCTCTCTGTGAACAGGGGCGCATTTGTGGCGCTGACACACCCCTGTGCGTGGGACTATCA 799

```

QY		721	CAGAAGCATTACGACATCATCCCTGACGAAGTGGCCCGAGCTGGGGGGAGATGCAC	780
Dd		800	CAAGAACATTTCACGATCATCCCTGACGAAGTGGCCCGAGCTGGGGGGAGATGCAC	859
QY		781	AATAAAGAAAGATGCTGAGCTTTCTGGTGSCCTTTGGCTCTAATTCTGGCCACACAGAAT	840
Dd		860	AATAAAGAAAGATGCTGAGCTTTCTGGTGSCCTTTGGCTCTAATTCTGGCCACACAGAAT	919
QY		841	CAGTGGCCCTATTGTGCAACTCTTTGTTCCCGGGACCTTGGTGGAGTTCTGACGCCAC	900
Dd		920	CAGTGGCCCTATTGTGCAACTCTTTGTTCCCGGGACCTTGGTGGAGTTCTGACGCCAC	979
QY		901	AAGAGCATGSGCGATTTTGAACGCCACATACCCTGTGGTGCACATGCCATACGCTGGTT	960
Dd		980	AAGAGCATGSGCGATTTTGAACGCCACATACCCTGTGGTGCACATGCCATACGCTGGTT	1033
QY		961	GACGGAATTCCTGTCTGTATCTCTGCTTTGCAAGATTAACCCGACACAACCTGCT	1020
Dd		1040	GACGGAATTCCTGTCTGTATCTCTGCTTTGCAAGATTAACCCGACACAACCTGCT	1099
QY		1021	GAGCACGAGCAACACAGTCAAGGAGGCCACATCAAGAGTTCACTCTGTCTGTGATGGC	1088
Dd		1100	GAGCACGAGCAACACAGTCAAGGAGGCCACATCAAGAGTTCACTCTGTCTGTGATGGC	1159
QY		1081	TCTGTCACTACGCTCTGTGTTCTGTATGTTTGAACACCCCAACGCTGTCTGAAAAATCTT	1140
Dd		1160	TCTGTCACTACGCTCTGTGTTCTGTATGTTTGAACACCCCAACGCTGTCTGAAAAATCTT	1219
QY		1141	GATPACACATCATCGGAGTGGANATTGGCCTTTGCAACATCTGTGTGTTCTTTGCGGAT	1200
Dd		1220	GATPACACATCATCGGAGTGGANATTGGCCTTTGCAACATCTGTGTGTTCTTTGCGGAT	1279
QY		1201	CTTCTCTCTTCCAGTTCACAGTCAACAGTGAAGGCGCATCTCAACGGGTGGCTGATAC	1266
Dd		1280	CTTCTCTCTTCCAGTTCACAGTCAACAGTGAAGGCGCATCTCAACGGGTGGCTGATAC	1339
QY		1261	ACTGAAGAAAACCTTCTGTCTGTTGCCCGGACCTGTGCTGGGATCATCGTCTCATAGC	1320
Dd		1340	ACTGAAGAAAACCTTCTGTCTGTTGCCCGGACCTGTGCTGGGATCATCGTCTCATAGC	1399
QY		1321	CAGCTCGTGTCCCAACCTTACCTTGGGGGTGACGCTGCGACCTGGGCGTGGGCTCCT	1380
Dd		1400	CAGCTCGTGTCCCAACCTTACCTTGGGGGTGACGCTGCGACCTGGGCGTGGGCTCCT	1459
QY		1381	CCTGGCGGGCTTGTGGGAGATTCACACATGGTGGCCATCGCTGCGTCAATGCTCAACG	1440
Dd		1460	CCTGGCGGGCTTGTGGGAGATTCACACATGGTGGCCATCGCTGCGTCAATGCTCAACG	1519
QY		1441	GAAGCAGAAAAAGAAAGATGAGAAATGAGTGGCCACCGAGGGGGGAGAACTCTGCCATAC	1500
Dd		1520	GAAGCAGAAAAAGAAAGATGAGAAATGAGTGGCCACCGAGGGGGGAGAACTCTGCCATAC	1579
QY		1501	AGACATGCTCTCGACAGAGAGGTGACAGACATCTGGAATTAGACAGAGAAATGAATA	1560
Dd		1580	AGACATGCTCTCGACAGAGAGGTGACAGACATCTGGAATTAGACAGAGAAATGAATA	1639
QY		1561	AGGACCGGAGACGCAATGGGCACTGACAGGAGCGGTAGTCAAGATGACACTTGGGACAT	1620
Dd		1640	AGGACCGGAGACGCAATGGGCACTGACAGGAGCGGTAGTCAAGATGACACTTGGGACAT	1699
QY		1621	CTCTTCCCTCCCATCGATTTTGTCCCTTTTGTGTTGTTTGTGTAAGAAGA	1680
Dd		1700	CTCTTCCCTCCCATCGATTTTGTCCCTTTTGTGTTGTTTGTGTAAGAAGA	1759
QY		1681	GGCCGTATTTAAAGTTTGGTGTGTCATTTCTCTAGACATCTGGGTATGCTACCTAGC	1740
Dd		1760	GGCCGTATTTAAAGTTTGGTGTGTCATTTCTCTAGACATCTGGGTATGCTACCTAGC	1819
QY		1741	GGGGACCTAGTGAATGGTCTTTTACTGTGTGCTAATGTAATAACAAACGAAACAACTGACTT	1800
Dd		1820	GGGGACCTAGTGAATGGTCTTTTACTGTGTGCTAATGTAATAACAAACGAAACAACTGACTT	1879
		1801	CATACCCCTGCTTACGAAAAACCAAAGACACAGCTGCTCACAGTTGACGTTGTGCTC	1860

D	b	1880	CATACCCCTGCTCACGAAAACCCAAAAGACACAGCTGCTTCACGTTGAAGTTGTGTC	1935
Q	y	1861	TCCTCCCTCGTGAACAATCTCTCTTGTGAAACCAAAGACTGAGCTGTGCATCGACCTCG	1920
D	b	1940	TCCTCCCTCGTGAACAATCTCTCTTGTGAAACCAAAGACTGAGCTGTGCATCGACCTCG	1995
Q	y	1921	GTACCCCTGCAACAGACGCCACAGACTCTCTGTCTCCCTTCATGTCTTTAAGATCAA	1980
D	b	2000	GTACCCCTGCAACAGACGCCACAGACTCTCTGTCTCCCTTCATGTCTTTAAGATCAA	2005
Q	y	1981	CAGTTTAAACCTGGCTTCCTTTGATTTGCTTCCAGTCAACAATGCCCTTACAAAGATG	204
D	b	2060	CAGTTTAAACCTGGCTTCCTTTGATTTGCTTCCAGTCAACAATGCCCTTACAAAGATG	2119
Q	y	2041	GAAGCCCGGTGGCTCTTAAATTTCCCTTGTGCAAGAGTTTCGAAACCATCTACAC	2100
D	b	2120	GAAGCCCGGTGGCTCTTAAATTTCCCTTGTGCAAGAGTTTCGAAACCATCTACAC	2175
Q	y	2101	ACATGACAGAGCGGGGTGGCAAGCTGCAAGCCCGAGTCCCGTTTCAACTGAGAACGGA	2166
D	b	2180	ACATGACAGAGCGGGGTGGCAAGCTGCAAGCCCGAGTCCCGTTTCAACTGAGAACGGA	2233
Q	y	2161	GACCTGTACACACAGCAGCTGACAGATGACAGAACTCCCGTAGAAGATTGGTTTG	2220
D	b	2240	GACCTGTACACACAGCAGCTGACAGATGACAGAACTCCCGTAGAAGATTGGTTTG	2299
Q	y	2221	AAATGCCCGGAGGAGCAAACTGACATGATTTGATGATGATGATCTCTGGCTTCTC	2280
D	b	2300	AAATGCCCGGAGGAGCAAACTGACATGATTTGATGATGATGATCTCTGGCTTCTC	2359
Q	y	2281	CTAGATCTGAGCAAGCTGTCAAGTTCTCACCCCTCACCGTGTATATACATGAGCTAACCTTT	2340
D	b	2360	CTAGATCTGAGCAAGCTGTCAAGTTCTCACCCCTCACCGTGTATATACATGAGCTAACCTTT	2419
Q	y	2341	TTAAATGTGCAAAAAGGCATCTCCAGATTCCAGACCCCGGCGCATGACCTTTCCGAA	2400
D	b	2420	TTAAATGTGCAAAAAGGCATCTCCAGATTCCAGACCCCGGCGCATGACCTTTCCGAA	2479
Q	y	2401	GCGTGTCTTTCCCTGCGCTTCTGGAAGTGCATTTAGACGAGTCAATGAGCATCC	2460
D	b	2480	GCGTGTCTTTCCCTGCGCTTCTGGAAGTGCATTTAGACGAGTCAATGAGCATCC	2539
Q	y	2461	TAACTTGCATTTTAACTTTTACAGTGAACCTGAAGCTTAACTCATCCAGCATTTCTAA	2520
D	b	2540	TAACTTGCATTTTAACTTTTACAGTGAACCTGAAGCTTAACTCATCCAGCATTTCTAA	2599
Q	y	2521	TGCCAGGTGCTGAGGGTAACTTTGAAGTATATATACCTGGTCTGCTATCCCTTA	2580
D	b	2600	TGCCAGGTGCTGAGGGTAACTTTGAAGTATATATACCTGGTCTGCTATCCCTTA	2659
Q	y	2581	GTCAATTAATCTGGCGGTACAGGTAATGGAATGTACTACGGTACTTCCCTCCACACAT	2640
D	b	2660	GTCAATTAATCTGGCGGTACAGGTAATGGAATGTACTACGGTACTTCCCTCCACACAT	2719
Q	y	2641	ACGATTAAGCAAGCATTTTATACGATACCAAGTCACTATGTGTCTTCCCTGAATA	2700
D	b	2720	ACGATTAAGCAAGCATTTTATACGATACCAAGTCACTATGTGTCTTCCCTGAATA	2779
Q	y	2701	ACGATTCGAATCCATGACAGTGCAGATATTTTCTAAGTTTGGAAAGAGGTTTTTT	2760
D	b	2780	ACGATTCGAATCCATGACAGTGCAGATATTTTCTAAGTTTGGAAAGAGGTTTTTT	2839
Q	y	2761	CTTTTAAAAAATTTATAGACAGCGGTCACTAAATTTGATAGTCAGAAATTCCTAGCTGA	2820
D	b	2840	CTTTTAAAAAATTTATAGACAGCGGTCACTAAATTTGATAGTCAGAAATTCCTAGCTGA	2899
Q	y	2821	AAGAACTTAACCAAAAAATTTTAAAGATATATATATGCTGTATATGTTATGTAAT	2880
D	b	2900	AAGAACTTAACCAAAAAATTTTAAAGATATATATATGCTGTATATGTTATGTAAT	2959
Q	y	2881	TTATTTTAAAGCTATATATACATTTCCATTTTGGCATTTTCAATATAAATGTCTTAATACA	2940

Db 2960 TTAATTTAGGCTAATATACATTCTCTATTTTCGATTTTCATATTAATGCTTAATACA 3013

QY 2941 AAA 2943

Db 3020 ATA 3022

RESULT 4  
LOCUS BC014526

DEFINITION Homo sapiens ankylosis, progressive homolog (mouse), mRNA (CDNA clone MGC:11142 IMAGE:3837372), complete cds.

ACCESSION BC014526

VERSION BC014526.2 GI:3878662

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3013)  
Strauberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.U., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE JOURNAL MEDLINE 22388257 PUBMED 12477932

REFERENCE 2 (bases 1 to 3013)  
Strauberg, R.  
Direct Submission  
Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NCI-MGC Project URL: http://mgc.nci.nih.gov  
On Aug 19, 2003 this sequence version replaced gi:15778895.  
Contact: MGC help desk  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbio.org  
contact: amadan@systemsbio.org  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

FEATURES  
SOURCE  
Location/Qualifiers  
1..3013  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

gene  
/clone="MGC:11142 IMAGE:3837372"  
/issue\_type="Ovary, adenocarcinoma"  
/clone\_id="N1H\_MGC\_9"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1..3013  
/gene="ANKH"  
/note="synonyms: ANK, HANK, MANK, CMDJ, CCAL2, CEPBD"  
/db\_xref="locusid:56172"  
/db\_xref="MIM:605145"  
57..1535  
/codon\_start=1  
/product="ankylosis, progressive homolog"  
/protein\_id="AAH14526.1"  
/db\_xref="GI:15778896"  
/translation="NFKPPLTHWPLIFLVLGITNTAIDPQNALNRGIAVAKED  
AVEMLSYGLAYELKPEFPMDFPMDFVFNRSKDRKVALCMVAALAVPFT  
LIVASPLGYIINKLHHVDESFGSKRRALVLAAPPMDAMATGILLKHSPL  
VGCASISDYIAQVFPVAILLHSHLECRPLITPLSYMALVRCITLCLGYKNHD  
IIPDRGPELGGATIRKMLSPWPLALILATORISRPVNLVFSRDGSSANTEAV  
AIIATAYPVGVHMYGMILTEIRAVYPAFADKNNPNSKLVSTNTVTAHITKKEFVCMAL  
SLTLCVMEWTPVNSKILIDIGVDFAEALCVPLRIKPSFPPVPTVAHITGWL  
TLKCTFLAPSSVLRILVLAISLVVPLVGHGATIGVGLAGFVGESTVMAIACY  
VVRKKKKKNNESATGEBDSAMTDMPTREVDIVEMRENE"

ORIGIN  
Query Match 98.4%; Score 2896.8; DB 9; Length 3013;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2912; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 27 CCGCAGAGTCCCTCGGCGAGCAGATGTGTGGGGTCAAGCCACGGCGGGGACTATG 86  
Db 1 CCGCAGAGTCCCTCGGCGAGCAGATGTGTGGGGTCAAGCCACGGCGGGGACTATG 60

QY 87 TGAATATCCCGCGCTCAGCAGCACTAGTGGCCCTGATCCGATCTTGTGTCCTCGGCA 146  
Db 61 TGAATATCCCGCGCTCAGCAGCACTAGTGGCCCTGATCCGATCTTGTGTCCTCGGCA 120

QY 147 TCACCAATAGCCATGACCTTGGGGAGAGAGGCTTGAACCGGGGCAATGCTGCTGCA 206  
Db 121 TCACCAATAGCCATGACCTTGGGGAGAGAGGCTTGAACCGGGGCAATGCTGCTGCA 180

QY 207 AGAGAGATGAGTGAATAGTGTGGCCAGCTAGCGGGCTGGTACTCCCTCATGAATCT 266  
Db 181 AGAGAGATGAGTGAATAGTGTGGCCAGCTAGCGGGCTGGTACTCCCTCATGAATCT 240

QY 267 TCACGGGTCCCATGAGTGAATGAGTGAATGAGGCTGCTGTTGTGAACAGCAAGAG 326  
Db 241 TCACGGGTCCCATGAGTGAATGAGTGAATGAGGCTGCTGTTGTGAACAGCAAGAG 300

QY 327 ACAGGACCAAGCGCTGCTGTATGTGTGTGGAGGAGGAGCCATGCTCGGCTTTTACA 386  
Db 301 ACAGGACCAAGCGCTGCTGTATGTGTGTGGAGGAGGAGCCATGCTCGGCTTTTACA 360

QY 387 CACGATAGCTTATAGTATGATTTAGATATCAATATCAATATCAATCAATGAGAG 446  
Db 361 CACGATAGCTTATAGTATGATTTAGATATCAATATCAATATCAATCAATGAGAG 420

QY 447 AGTGGTGGGAGAGCAAGAGAGAGGCTTCTGCTGCTGAGCGGCTTCTCTTCAATG 506  
Db 421 AGTGGTGGGAGAGCAAGAGAGGCTTCTGCTGCTGAGCGGCTTCTCTTCAATG 480

QY 507 ACGCAATGAGTGAAGCCATGCTGAGCTTCTTAAACCAATATCAATGATTTCTGTG 566  
Db 481 ACGCAATGAGTGAAGCCATGCTGAGCTTCTTAAACCAATATCAATGATTTCTGTG 540

QY 567 GATGTGCTCAATTCAGATGATATAGCTCAGATGTTTGTGACATTTGCTTACA 626  
Db 541 GATGTGCTCAATTCAGATGATATAGCTCAGATGTTTGTGACATTTGCTTACA 600

QY 627 GTCACTGGAATGCGGAGAGCCCTGCTCATCCGATCTCTCTTGAATGAGGAGC 686